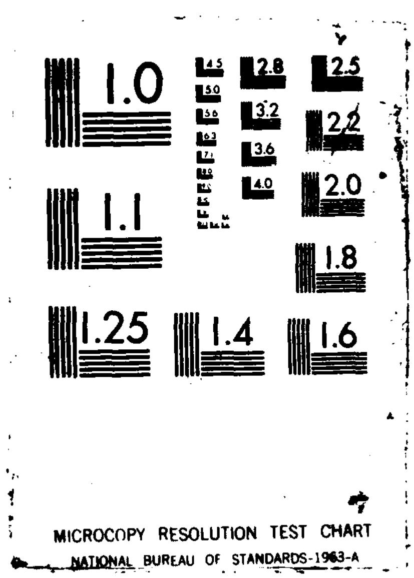
AD-A185 695 STRONG CONSISTENCY AND EXPONENTIAL RATE OF THE 'MINIMUM L1-NORM' ESTIMATE..(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS Y WU JUN 87 87-18 1/1

UNCLASSIFIED AFOSR-TR-87-0976 F49620-85-C-0008 F/G 12/3 NL


MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Unclassified

DTIC FILE COPY

2

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD-A185 695

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 87-0976	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Strong consistency and exponential rate of the "minimum L_1-norm" estimates in linear regression models		5. TYPE OF REPORT & PERIOD COVERED ~~Technical~~ Journal - June 1987
		6. PERFORMING ORG. REPORT NUMBER 87-18
7. AUTHOR(s) Yuehua Wu		8. CONTRACT OR GRANT NUMBER(s) F49620-85-C-0008
PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis Fifth Floor Thackeray Hall University of Pittsburgh, Pittsburgh, PA 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304 A5
CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research NM Department of the Air Force Bolling Air Force Base, DC 20332 Bldg 410		12. REPORT DATE June 1987
		13. NUMBER OF PAGES 25
MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) same as 11		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTE OCT 0 1 1987 S D

D

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

L_1-norm estimation, consistency, exponential rate, linear model, median regression.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let $y_i = \alpha + x_i'\underset{\sim}{\beta} + e_i$ $i = 1, \ldots, n, \ldots$ be a linear regression model, where $\{x_i\}$ is a sequence of experimental points, i. e., known p-vectors, $\{e_i\}$ is a sequence of independent random errors, with $\mathrm{med}(e_i) = 0$, $i = 1, 2, \ldots$ Define the minimum L_1-norm estimate of $(\alpha, \underset{\sim}{\beta}')'$, by $(\hat{\alpha}_n, \underset{\sim}{\hat{\beta}}_n')'$, to be chosen such that

DD FORM 1473, 1 JAN 73

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. (continued)

$$\sum_{i=1}^{n} |y_i - \hat{\alpha}_n - x_i'\hat{\beta}_n|$$

$$= \min_{\alpha, \beta} \sum_{i=1}^{n} |y_i - \alpha - x_i'\beta|.$$

Under quite general conditions on $\{x_i\}$ and $\{e_i\}$, the strong consistency of the minimum L_1-norm estimate is established. Further, under an additional condition on $\{x_i\}$, it is also proved that for any given $\varepsilon > 0$, there exist constant $C > 0$ not depending on n, such that

$$P\{||\hat{\alpha}_n - \alpha||^2 + ||\hat{\beta}_n - \beta||^2 \geq \varepsilon^2\}$$

$$\leq \exp\{-Cn\}, \qquad \text{for large } n.$$

AFOSR-TR- 87-0976

Center for Multivariate Analysis

University of Pittsburgh

STRONG CONSISTENCY AND EXPONENTIAL RATE OF THE "MINIMUM L_1-NORM" ESTIMATES IN LINEAR REGRESSION MODELS *

Yuehua Wu

Center for Multivariate Analysis
University of Pittsburgh

June 1987

Technical Report No. 87-18

Center for Multivariate Analysis
Fifth Floor Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

* This work is partially sponsored by Contract F49620-85-C-0008 of the Air Force Office of Scientific Research. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

Abstract

Let $y_i = \alpha + x_i'\underset{\sim}{\beta} + e_i$, $i = 1, \ldots, n, \ldots$ be a linear regression model, where $\{x_i\}$ is a sequence of experimental points, i. e., known p-vectors, $\{e_i\}$ is a sequence of independent random errors, with $\text{med}(e_i) = 0$, $i = 1, 2, \ldots$. Define the minimum L_1-norm estimate of $(\alpha, \beta')'$, by $(\hat{\alpha}_n, \hat{\beta}_n')'$, to be chosen such that

$$\sum_{i=1}^{n} |y_i - \hat{\alpha}_n - x_i'\underset{\sim}{\hat{\beta}}_n|$$
$$= \min_{\alpha, \underset{\sim}{\beta}} \sum_{i=1}^{n} |y_i - \alpha - x_i'\underset{\sim}{\beta}|.$$

Under quite general conditions on $\{x_i\}$ and $\{e_i\}$, the strong consistency of the minimum L_1-norm estimate is established. Further, under an additional condition on $\{x_i\}$, it is also proved that for any given $\varepsilon > 0$, there exist constant $C > 0$ not depending on n, such that

$$P\{||\hat{\alpha}_n - \alpha||^2 + ||\underset{\sim}{\hat{\beta}}_n - \underset{\sim}{\beta}||^2 \geq \varepsilon^2\}$$
$$\leq \exp\{-Cn\}, \qquad \text{for large } n.$$

AMS 1980 subject classifications: Primary 62J05; Secondary 62F12.

Key words and phrases: L_1-norm estimation, consistency, exponential rate, linear model, median regression.

1. Introduction.

Consider the linear regression model

$$y_i = \alpha + \underset{\sim}{x}_i'\underset{\sim}{\beta} + e_i, \qquad i = 1, 2, \ldots \tag{1.1}$$

where $\{ \underset{\sim}{x}_i = (x_{i1}, \ldots, x_{ip})' \}$, $i = 1, 2, \ldots$, is a sequence of experiment points, i. e. known p-vectors, $\underset{\sim}{\beta} = (\beta_1, \ldots, \beta_p)'$ is the regression-coefficient vector, and $e_1, e_2, \ldots$ are random errors. To estimate the unknown vector $\underset{\sim}{\beta}$ based on the observations $y_1, \ldots, y_n$, a popular method is the Least Squares (LS) method, which takes $\tilde{\underset{\sim}{\beta}}_n$ minimizing the sum of squares $\sum_{i=1}^{n} (y_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})^2$ as the estimate of $\underset{\sim}{\beta}$.

The merit of LS estimate is that it is easy to compute (being a linear combination of $y_1, \ldots, y_n$), and that, in case where the random errors are independently and approximately normally distributed, the LS estimate possesses many desirable properties. But these happy conditions cannot be taken for granted in many practical applications. For example, in econometrics, there now exists a considerable body of evidence that attests to distributions with infinite variance being a reality (distribution of income, behavior of speculative prices, distribution of firms by size etc.). Even in cases where the error variance can reasonably be assumed to be finite, the error distributions may have heavy tails, deteriorating the performance of the LS estimate.

So it is under this background there has now been much interest in using more robust methods, among which the L_1-norm method is a forerunner. This method seeks to minimize $\sum_{i=1}^{n} |y_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}|$ instead of $\sum_{i=1}^{n} (y_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})^2$, and the

minimizer $\hat{\underset{\sim}{\beta}}_n$ is taken as the estimate of $\underset{\sim}{\beta}$. True, the calculation of $\hat{\underset{\sim}{\beta}}_n$ is considerable more complicated as compared with the case of LS estimate $\tilde{\underset{\sim}{\beta}}_n$, but, since the successful establishment of the link between the L_1-norm method and the linear programming, the computing problem no longer presents a barrior in the face of modern computing facilities.

Another problem is the sampling theory of the estimate. Since it is not likely that a workable small-sample theory can be established, the asympototic theory is of great importance. The first and foremost problem in the asympototic theory is to establish the consistency of this estimate under weak conditions.

The asymptotic theory of the L_1-norm estimate is much more difficult as compared with the Least Squares theory, owing to the mathematical difficulty of working with the absolute value function. Huber (1981) proved a general theorem concerning the consistency and asymptotic normality of a class of robust estimates of linear regression coefficients, but the result does not apply to the L_1-norm estimate, since the absolute value function is not differentiable at zero. In recent years some authors, for example Oberhofer (1982), proved the weak consistency of the L_1-norm estimate under rather strong conditions.

In this paper, we develop a method in dealing with the consistency problem of the L_1-norm estimate which enables us to obtain general results concerning the strong consistency and exponential rate of the L_1-norm estimate under very mild conditions.

2. Formulation of the Theorems.

In some applications it may be known in advance that the regression plane passes through the origin, i. e. $\alpha = 0$ in model (1.1). In this case, according to the L_1-norm criterion, the solution $\hat{\underset{\sim}{\beta}}_n$ of the minimization problem

$$\sum_{i=1}^{n} |y_i - x_i'\hat{\beta}_n| = \min_{\beta} \sum_{i=1}^{n} |y_i - x_i'\beta| .$$

gives an estimation of β. Define

$$S_n = \sum_{i=1}^{n} x_i x_i' ,$$

$$\rho_n = \text{the smallest eigenvalue of } S_n ,$$

$$d_n = \max \{ 1, ||x_1||, \ldots, ||x_n|| \} ,$$

where $||x||$ denotes the Euclidean norm of vector x.

Theorem 1. Suppose the the following conditions are satisfied:

1) $\rho_n/(d_n^2 \log n) \to \infty$, (2.1)

2) There exists constant $k > 1$ such that

$d_n/n^{k-1} \to 0$, (2.2)

3) $e_1, e_2, \ldots$ are independent random variables, and

$\mathrm{med}(e_i) = 0, \qquad i = 1, 2, \ldots$

4) There exist constants $C_1 > 0$, $C_2 > 0$ such that

$$P\{\, -h < e_i < 0 \,\} \geq C_2 h, \tag{2.3}$$

$$P\{\, 0 < e_i < h \,\} \geq C_2 h, \tag{2.4}$$

for all $i = 1, 2, \ldots$ and $h \in (0\ ,\ C_1)$. Then we have

$$\lim_{n\to\infty} \hat{\underset{\sim}{\beta}}_n = \underset{\sim}{\beta}, \qquad \text{a. s.}$$

Further, under the additional condition that for some constant $M > 0$

$$\rho_n / d_n^2 \geq Mn, \qquad \text{for large } n \tag{2.5}$$

$\hat{\underset{\sim}{\beta}}_n$ converges to $\underset{\sim}{\beta}$ at an exponential rate in the following sense: For arbitrary given $\varepsilon > 0$ there exists constant $C > 0$ independent of n such that

$$P\{||\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}|| \geq \varepsilon\}$$

$$\leq O(e^{-Cn}).$$

Theoretically speaking, the nonhomogeneous model (1.1) is merely a special case of the homogeneous model $y_i = \underset{\sim}{x}_i'\beta + e_i$, $i = 1, 2, \ldots$, in which first element of each x_i is 1. Therefore, as a corollary, from Theorem 1 we can obtain an analogous result concerning the estimates $\hat{\alpha}_n$, $\hat{\underset{\sim}{\beta}}_n$ of α, $\underset{\sim}{\beta}$: We need only to replace the matrix $S_n = \sum_{i=1}^{n} x_i x_i'$ by

$$\tilde{S}_n = \begin{pmatrix} n & \sum_{i=1}^{n} x_i' \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i x_i' \end{pmatrix}$$

and define ρ_n as the smallest eigenvalue of $\tilde{S}_n$. But, since $\tilde{S}_n$ is a matrix of higher order as compared with S_n, it may be of some advantage to give the following result

Theorem 2. Suppose that the we have model (1.1), and the conditions of Theorem 1 are satisfied, except that here we define S_n as $\sum_{i=1}^{n} (x_i - \bar{x}_n)(x_i - \bar{x}_n)'$ where $\bar{x}_n = (1/n) \sum_{i=1}^{n} x_i$, then

$$\lim_{n\to\infty} \hat{\alpha}_n = \alpha, \qquad \text{a. s.}$$

$$\lim_{n\to\infty} \hat{\underset{\sim}{\beta}}_n = \underset{\sim}{\beta}, \qquad \text{a. s.}$$

Also, under the additional assumption (2.5) for arbitrarily given $\varepsilon > 0$ we can find constant $C > 0$ independent of n such that

$$P\{||\hat{\alpha}_n - \alpha||^2 + ||\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}||^2 \geq \varepsilon^2\}$$

$$\leq O(e^{-Cn}).$$

In this formulation, Theorem 2 is not a trival consequence of Theorem 1, and

through the basic idea of proof are the same for the two theorems, some important differences in detail emerge. Therefore we shall give the detailed proof of both.

3. Comments on Conditions.

Before entering the details of proofs, we make some remarks concerning the conditions of the theorems.

1. The first two conditions involve only the sequence of experiment points $\{x_i\}$, while the latter ones involve only the error sequence $\{e_i\}$. In any theoretical problem concerning the linear model (1.1), it is always desirable not to introduce assumptions involving both simultaneously.

2. Conditions 3) and 4), taken together, guarantee the uniqueness of the median of e_i, $i = 1, 2, \ldots$. Condition 4) stipulates that the random errors (e_i's) should not be "too lightly" distributed around the median zero. The requirement on the uniqueness of median is reasonable in view of the "median-regression" character of the model. As for the conditions (2.3) and (2.4), it is likely that they are not necessary and further improvements are conceivable, yet it is easily seen that they cannot be totally dispensed with.

Example 1. Take the simplest case in which we know in advance that $\beta = 0$. In this case the Minimum L_1-norm principle gives

$$\hat{\alpha}_n = \text{med}\,(y_1, \ldots, y_n)$$

as the estimate of α. Suppose that $e_1, e_2, \ldots$ are mutually independent, e_i has the following density function:

$$f_i(x) = \begin{cases} |x|/i^2, & 0 \le |x| \le i, \\ 0, & \text{otherwise}, \end{cases} \qquad i = 1, 2, \ldots$$

Then

$$P\{e_i \ge 1\} = 1/2 - 1/(2i^2), \qquad i = 1, 2, \ldots$$

Denote by ξ_n the number of those e_i's for which $\sqrt{n} \le i \le n$ and $e_i \ge 1$. An application of the Central Limit Theorem convinces us that for given $\delta \in (0, 1/2)$ we have

$$P\{\xi_n > n/2\} \ge \delta$$

for n sufficiently large. This implies that

$$P\{\hat{\alpha}_n \ge 1\} \ge \delta$$

for n sufficiently large, hence $\hat{\alpha}_n$ is not a consistent estimate of α.

3. Conditions 1) and 2) regulate the behavior of the sequence of experiment points $\{x_n\}$. 2) stipulates simply that x_n should not go to infinity "too fast" as $n \to \infty$. A close inspection of condition 1) convinces us that this is also implied by 1).

Condition 1) requires that ρ_n should tend to infinity with some rate. In the case where $\{x_i\}$ is bounded, ρ_n should tend to infinity at a rate faster than logn. In the LS theory, under some general conditions (see Li(1984)) on the error sequence $\{e_i\}$, the strong consistency of LS estimates is guaranteed by requireing only $\rho_n \to \infty$. This gives one the hope that condition 1) can be weakened to $\rho_n \to \infty$. Whether or not this is true remains an open question.

The point that "x_n should not go to infinity too fast" is of interest as it

reveals a difference between the L_1-norm and LS criteria. For LS estimate, in general the faster x_n goes to infinity as $n \to \infty$, the more likely is that it becomes consistent. The following example shows that in the L_1-norm case, going too fast (of x_n) to infinity may indeed render the estimate inconsistent.

Example 2. Suppose that in model (1.1) $p = 1$, the random errors e_1, $e_2, \ldots$ independent, $P\{e_i = 10^i\} = P\{e_i = -10^i\} = 1/6$, and e_i is uniformly distributed over the interval $(-1/3, 1/3)$ with density one. For convenience assume that the true parameter values are $\alpha = 0$, $\underset{\sim}{\beta} = 0$. Let $x_i = 10^i$, $i = 1, 2, \ldots$.

Define the event $E_n = \{e_n = 10^n\}$. Then $P\{E_n\} = 1/6$. It is readily seen that when E_n occurs we have

$$\sum_{i=1}^{n} |e_i - a - bx_i| \geq 4\times10^{n-1}, \text{ when } |a| < 1/10 \text{ and } |b| < 1/10,$$

and

$$\sum_{i=1}^{n} |e_i - a - bx_i|_{a=0,b=1} \leq 3\times10^{n-1}.$$

These two facts, taken together, give

$$P\{\max\ (|\hat{\alpha}_n|, |\hat{\beta}_n|) \geq 1/10\} \geq P\{e_n\} = 1/6.$$

So even weak consistency does not hold.

Example 3. This example shows that, even in the case that $e_1, e_2, \ldots$ are independent and identically distributed, consistency may not hold in case p_n tends to infinity too fast.

Suppose that in model (1.1) $p=1$, the true parameters $\alpha = 0$, $\beta = 0$, the random errors are iid. with a common distribution $P\{e_1 = 10^k\} = P\{e_1 = -10^k\} =$

$1/[k(k+1)]$, $k = 6, 7, \ldots$ and e_1 is uniformly distributed over $(-1/3, 1/3)$ with density 1. Let $x_i = 10^i$, $i = 1, 2, \ldots$

Define the event $E_n = \{|e_i| < 10^n, n/2 < i \le n\}$. Then it is readily seen that $P(E_n) \to e^{-2}$ as $n \to \infty$. Suppose that E_n does not occur and denote by i_n the first j such that $j > n/2$ and $|e_j| \ge 10^n$, then we have $i_n \le n$. An argument similar to that employed in Example 2 gives us that $\max(|\hat{\alpha}_{i_n}|, |\hat{\beta}_{i_n}|) \ge 1/10$. Therefore $P\{|\hat{\alpha}_i| \ge 1/10$ or $|\hat{\beta}_i| \ge 1/10$, for some i, $n/2 < i \le n\} \ge P\{E_n^c\} \to 1 - e^{-2} > 0$ and $(\hat{\alpha}_n, \hat{\beta}_n)$ is not strongly consistent.

From a practical point of view it can be said that the conditions 1) and 2) are reasonable and would be satisfied in most applications. An important case is that $\{x_i\}$ is bounded, or more generally, $||x_i|| = o(i/\log i)$, and

$$\lim_{n\to\infty} S_n/n = \Lambda \text{ exists and positive definite.} \tag{3.1}$$

These conditions are satisfied when $x_1, x_2, \ldots$ are iid. samples of a random vector x for which $E(xx')$ exists and is positive definite.

4. Proof of Theorem 1.

The following lemmas will be needed in the proof.

Lemma 1 (Bennett.) Suppose that $\xi_1, \ldots, \xi_n$ are independent random variables with $|\xi_i| \le b < \infty$. Let $\mu = (1/n)\sum_{i=1}^{n} E\xi_i$ and $\sigma^2 = (1/n)\sum_{i=1}^{n} \mathrm{Var}\,\xi_i$. Then for each $\varepsilon > 0$, we have

$$P\{|(1/n)\sum_{i=1}^{n} \xi_i - \mu| \geq \varepsilon\}$$

$$\leq 2\exp\{-n\varepsilon^2/[2(\sigma^2 + b\varepsilon)]\}.$$

The proof of this lemma can be found in Hoeffiding (1963).

Lemma 2. Suppose that $\{e_i\}$, $i = 1, 2, \ldots, n$ are random variables satisfying the conditions 3) and 4) of Theorem 1. Also, let $\{a_i\}$, $i = 1, \ldots, n$ be a sequence of constants and $a^* = \max_{1\leq i\leq n} |a_i|$. Then there exist constants $C > 0$, and $\varepsilon_0 > 0$, depending only on a^*, C_1 and C_2, such that

$$P\{\sum_{i=1}^{n} (|e_i| - |e_i - a_i|) > -\varepsilon_0 B_n^2\}$$

$$\leq 2\exp\{-CB_n^2\}, \tag{4.1}$$

where $B_n^2 = \sum_{i=1}^{n} a_i^2$.

Proof. Without loss of generality, we can assume that $a_i \geq 0$, for each $i = 1, \ldots, n$. Note that (4.1) is automatically true if $a^* = 0$. Thus we can assume that $a^* > 0$, which implies that $B_n^2 > 0$. In case of $a^* > C_1$, we can suitably adjust C_2 such that (2.3) and (2.4) are true for each i and $h \varepsilon (0, a^*)$.

Define $\xi_i = |e_i| - |e_i - a_i|$. It is easy to see that

$$\xi_i = \begin{cases} a_i, & \text{if } e_i \geq a_i \\ -a_i, & \text{if } e_i \leq 0 \\ 2e_i - a_i, & \text{if } 0 < e_i < a_i. \end{cases}$$

Hence

$$|\xi_i| \leq a_i \leq a^*,$$

$$E\xi_i \leq -a_i P\{e_i \leq 0\} + a_i P\{e_i > a_i/2\},$$

$$\leq -a_i P\{0 < e_i \leq a_i/2\} \leq -C_2 a_i^2/2,$$

$$\mathrm{Var}\,\xi_i \leq E\xi_i^2 \leq a_i^2$$

Take $\varepsilon_0 = C_2/4$ and $C = \varepsilon_0^2/[2(1 + a^*\varepsilon_0)]$. Using Lemma 1, we have

$$P\{\sum_{i=1}^n \xi_i \geq -\varepsilon_0 B_n^2\}$$

$$= P\{(1/n)\sum_{i=1}^n (\xi_i - E\xi_i) \geq -(1/n)(\sum_{i=1}^n E\xi_i + \varepsilon_0 B_n^2)\}$$

$$\leq P\{(1/n)\sum_{i=1}^n (\xi_i - E\xi_i) \geq \varepsilon_0 B_n^2/n\}$$

$$\leq 2\exp\{-n(\varepsilon_0 B_n^2/n)^2/[2(B_n^2/n + a^*\varepsilon_0 B_n^2/n)]\}$$

$$= 2\exp\{-CB_n^2\},$$

which comletes the proof of Lemma 2.

Lemma 3. Maintaining the assumptions and notations given in Lemma 2, and defining

$$\eta_i = \begin{cases} 1, & \text{if } e_i > a_i/2 \\ -1, & \text{if } e_i \leq a_i/2, \end{cases}$$

we have

$$P\{\sum_{i=1}^{n} a_i\eta_i > -\varepsilon_0 B_n^2\} \leq 2\exp\{-CB_n^2\}.$$

The proof runs largely along the same line as in Lemma 2. So the details are omitted.

Now turn to the proof of the theorem. Without loss of generality we assume that the true parameters $\underset{\sim}{\beta} = 0$.

Fix ε, $0 < \varepsilon < 1$, and define $\Lambda = \{\underset{\sim}{\beta}: ||\underset{\sim}{\beta}|| = \varepsilon\}$. Split Λ into $m = m_n$ parts $\Lambda_1, \ldots, \Lambda_m$ such that the diameter of each part does not exceed ε/n^k. It is easy to see that this can be done with

$$m = m_n \leq (pn)^{pk}.$$

Choose arbitrarily a point $\underset{\sim}{\beta}_t$ from each Λ_t, $t = 1, \ldots, m$. Define $b_{ti} = x_i'\underset{\sim}{\beta}_t$, $b_t = \max_{1\leq i\leq n} \{|b_{ti}|, 1\}$ and $a_{ti} = b_{ti}/b_t$. Then we have $|a_{ti}| \leq 1$ for each t and i. Also, we have

$$B_{tn}^2 = \sum_{i=1}^{n} a_{ti}^2 = \underset{\sim}{\beta}_t' S_n \underset{\sim}{\beta}_t / b_t^2$$

$$\geq \rho_n/d_n^2 \to \infty.$$

Hence, by Lemma 2, we have, for large n,

$$P\{\sum_{i=1}^n (|e_i| - |e_i - a_{ti}|) > -\varepsilon\}$$

$$\leq P\{\sum_{i=1}^n (|e_i| - |e_i - a_{ti}|) \geq -\varepsilon B_{tn}^2\}$$

$$\leq 2\exp\{-CB_{tn}^2\} \leq 2\exp\{-C\varepsilon^2 \rho_n/d_n^2\}.$$

If $\underset{\sim}{\beta} \varepsilon \Lambda_t$ and $\sum_{i=1}^n (|e_i| - |e_i - a_{ti}|) \leq -\varepsilon$, then

$$\sum_{i=1}^n (|e_i| - |e_i - x_i'\underset{\sim}{\beta}/b_t|)$$

$$\leq \sum_{i=1}^n (|e_i| - |e_i - a_{ti}|) + \sum_{i=1}^n |x_i'(\underset{\sim}{\beta} - \underset{\sim}{\beta}_t)|/b_t$$

$$\leq -\varepsilon + nd_n\varepsilon/n^k < 0, \qquad \text{for all large } n.$$

Here the last step follows from condition (2.2). Therefore, for all large n we have

$$P\{\sum_{i=1}^n |e_i| - \min_{\underset{\sim}{\beta}\varepsilon\Lambda_t'} \sum_{i=1}^n |e_i - x_i'\underset{\sim}{\beta}| \geq 0\}$$

$$\leq P\{\sum_{i=1}^n (|e_i| - |e_i - a_{ti}|) > -\varepsilon\}$$

$$\leq 2\exp\{-C\varepsilon^2 \rho_n / d_n^2\} \tag{4.2}$$

where $\Lambda_t' = \{b_t^{-1}\underset{\sim}{\beta}\colon \underset{\sim}{\beta} \in \Lambda_t\}$

Now let $\ell > 1$. Define

$$\eta_{ti} = \begin{cases} 1, & \text{if } e_i > a_{ti}/2 \\ -1, & \text{if } e_i \leq a_{ti}/2 \end{cases} \qquad i = 1, \ldots, n$$

and

$$\Delta_i = |e_i| - |e_i - \ell a_i|,$$

where $a_i = x_i'\underset{\sim}{\beta}$, $\underset{\sim}{\beta} \in \Lambda_t'$. Let us consider the following cases. In each case, we shall use the fact that $|a_{ti} - a_i| = |x_i'(\underset{\sim}{\beta} - \underset{\sim}{\beta}_t)|/b_t \leq d_n\varepsilon/n^k = o(1/n)$. Here and in the sequel, $o(1/n)$ is uniform in t and i.

Case *a*. $a_i \geq a_{ti}/2 > 0$ and $e_i < a_{ti}/2$.

We have

$$\Delta_i = |e_i| - |e_i - a_i| - (\ell - 1)a_i$$

$$\leq |e_i| - |e_i - a_i| - (\ell - 1)(a_{ti} - o(1/n))$$

$$= |e_i| - |e_i - a_i| + (\ell - 1)(a_{ti}\eta_{ti} + o(1/n)) \tag{4.3}$$

Case *b*. $a_i \geq a_{ti}/2 > 0$, and $e_i \geq a_{ti}/2$.

We have

$$\Delta_i \le |e_i| - |e_i - a_i| + (\ell - 1)a_i$$

$$\le |e_i| - |e_i - a_i| + (\ell - 1)(a_{ti} + o(1/n))$$

$$= |e_i| - |e_i - a_i| + (\ell - 1)(a_{ti}\eta_{ti} + o(1/n)),$$

which shows that (4.3) is still true.

Case *c*. $0 > a_{ti}/2 \geq a_i$ and $e_i > a_{ti}/2$.

We have

$$\Delta_i = |e_i| - |e_i - a_i| + (\ell - 1)a_i$$

$$\le |e_i| - |e_i - a_i| + (\ell - 1)(a_{ti} + o(1/n))$$

$$= |e_i| - |e_i - a_i| + (\ell - 1)(a_{ti}\eta_{ti} + o(1/n)).$$

Hence (4.3) is still true for this case.

Case *d*. $0 > a_{ti}/2 \geq a_i$ and $e_i \leq a_{ti}/2$.

We have

$$\Delta_i \le |e_i| - |e_i - a_i| + (\ell - 1)|a_i|$$

$$\le |e_i| - |e_i - a_i| + (\ell - 1)(|a_{ti}| + o(1/n))$$

$$= |e_i| - |e_i - a_i| + (\ell - 1)(a_{ti}\eta_{ti} + o(1/n)),$$

which also shows that (4.3) is true.

Case *e*. $|a_i| \geq |a_{ti}|/2$.

In this case, we have that $|a_{ti}| \leq 2|a_{ti} - a_i| = o(1/n)$ hence $a_i = o(1/n)$. Therefore

$$\Delta_i = |e_i| - |e_i - \ell a_i|$$

$$\leq |e_i| - |e_i - a_i| + (\ell - 1)o(1/n)$$

$$= |e_i| - |e_i - a_i| + (\ell - 1)(a_{ti}\eta_{ti} + o(1/n)),$$

and (4.3) is still true.

Summing up the five cases, we see that

$$\Delta_i \leq |e_i| - |e_i - a_i| + (\ell - 1)(a_{ti}\eta_{ti} + o(1/n))$$

Let $\Pi'_t = \{\ell\beta: \beta \in \Lambda'_t \text{ and } \ell \geq 1\}$. By (4.2), (4.3) and Lemma 3, we have

$$P\{\sum_{i=1}^n |e_i| - \min_{\beta \in \Pi'_t} \sum_{i=1}^n |e_i - x_i'\beta| \geq 0\}$$

$$\leq P\{\sum_{i=1}^n |e_i| - \min_{\beta \in \Lambda'_t} \sum_{i=1}^n |e_i - x_i'\beta| \geq 0\}$$

$$+ P\{\sum_{i=1}^n a_{ti}\eta_{ti} \geq -\varepsilon\}$$

$$\leq 4\exp\{-C\varepsilon^2\rho_n/d_n^2\}.$$

Denote $\Pi_t = \{\ell\beta: \beta \in \Lambda_t \text{ and } \ell \geq 1\}$. Note that $b_t \geq 1$, we have $\Pi_t \subset \Pi'_t$.

Also, $\{||\beta|| \geq \varepsilon\} = \bigcup_{t=1}^m \Pi_t \subset \bigcup_{t=1}^m \Pi'_t$. Hence we obtain

$$P\{\sum_{i=1}^{n} |e_i| - \min_{||\underset{\sim}{\beta}||\geq\varepsilon} \sum_{i=1}^{n} |e_i - x_i'\underset{\sim}{\beta}| \geq 0\}$$

$$\leq \sum_{t=1}^{n} P\{\sum_{i=1}^{n} |e_i| - \min_{\underset{\sim}{\beta}\varepsilon\Pi_t'} \sum_{i=1}^{n} |e_i - x_i'\underset{\sim}{\beta}| \geq 0\}$$

$$\leq 4m\exp\{-C\varepsilon^2\rho_n/d_n^2\}$$

$$\leq 4(pn)^{pk}\exp\{-C\varepsilon^2\rho_n/d_n^2\} \tag{4.4}$$

which, together with (2.1) and Borel–Cantelli Lemma, implies that

$$P\{||\hat{\underset{\sim}{\beta}}_n|| \geq \varepsilon,\ i.\ o.\} = 0$$

This proves the strong consistency of $\hat{\underset{\sim}{\beta}}_n$. Finally, the last assertion of the theorem follows directly from (2.5) and (4.4). The proof is concluded.

Remark. If (3.1) holds, then condition (2.2) for $k = 3/2$ is fulfilled. If we further assume that $d_n = o(\sqrt{n/\log n})$, then (2.1) is satisfied. Also, it is easy to see that (3.1) and the boundedness of d_n imply (2.2) and (2.5) which guarantee the exponential rate of the convergence of the L_1-norm estimate. It may be noticed that since in general we have $\rho_n \leq nd_n^2$, we cannot get a rate faster than $O(e^{-Cn})$ from the proof of Theorem 1. In fact, under the assumption that $e_1, e_2, \ldots$ are independent and identically distributed, it can easily be shown that $P\{||\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}|| \geq \varepsilon\}$ cannot tend to zero at a rate faster than $O(e^{-Cn})$ (i. e. $P\{||\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}|| \geq \varepsilon\} = O(g(n))$ and $e^{Cn}g(n) \to 0$ for any $C > 0$ as $n \to \infty$).

5. Proof of Theorem 2.

Without loss of generality, we can assume that $\alpha = 0$, $\underset{\sim}{\beta} = \mathbf{0}$.

Given ε, $0 < \varepsilon < 1$, and define

$$\Lambda = \{(\alpha, \underset{\sim}{\beta}')' : \alpha^2 + \underset{\sim}{\beta}'\underset{\sim}{\beta} = \varepsilon^2\}.$$

Split Λ into $m = m_n$ parts $\Lambda_1, \ldots, \Lambda_m$ such that the diameter of each part does not exceed ε/n^k and that $m \le ((p+1)n)^{(p+1)k} = C_0 n^{(p+1)k}$ where C_0 is a positive constant. Choose arbitrarily a point $(\alpha_t, \underset{\sim}{\beta}'_t)'$ from Λ_t and define

$$b_{ti} = x_i'\underset{\sim}{\beta}_t, \qquad b_t = \max_{1 \le i \le n} |b_{ti}|.$$

Consider the following two cases.

Case 1. $b_t \le \varepsilon/4$.

Define $a_{ti} = \alpha_t + b_{ti}$. Then we have

$$|a_{ti}| \le \varepsilon + \varepsilon/4 \le 2, \qquad i = 1, \ldots, n.$$

Arguing as in the proof of Theorem 1, we get

$$P\{\sum_{i=1}^{n} (|e_i| - |e_i - a_{ti}|) > -\varepsilon\}$$

$$\leq 2\exp\{-C\sum_{i=1}^{n} a_{ti}^{2}\},$$

and consequently

$$P\{\sum_{i=1}^{n} |e_i| - \min_{\substack{(\alpha,\beta')'\varepsilon\Lambda_t \\ \ell\geq 1}} \sum_{i=1}^{n} |e_i - \ell(\alpha+x_i'\beta)| \geq 0\}$$

$$\leq 2\exp\{-C\sum_{i=1}^{n} a_{ti}^{2}\}. \tag{5.1}$$

Noticing that $b_t \leq \varepsilon/4$, we have

$$\sum_{i=1}^{n} a_{ti}^{2} = n(\alpha_t + \bar{x}_n'\beta_t)^2 + \beta_t' T_n \beta_t$$

$$\geq n(\varepsilon/2 - \varepsilon/4)^2 = n\varepsilon^2/16. \tag{5.2}$$

Hence, by (5.1) and (5.2) we have

$$P\{\sum_{i=1}^{n} |e_i| - \min_{\substack{(\alpha,\beta')'\varepsilon\Lambda_t \\ \ell\geq 1}} \sum_{i=1}^{n} |e_i - \ell(\alpha+x_i'\beta)| \geq 0\}$$

$$\leq 2\exp\{-Cn\}. \tag{5.3}$$

Case 2. $b_t > \varepsilon/4$.

Define $a_{ti} = \varepsilon(\alpha_t + b_{ti})/(4b_t)$. We have

$$|a_{ti}| \leq \varepsilon + \varepsilon/4 < 5/4.$$

As before, using Lemma 2, we can show that

$$P\{\sum_{i=1}^{n} (|e_i| - |e_i - a_{ti}|) > -\varepsilon\}$$

$$\leq 2\exp\{-C\sum_{i=1}^{n} a_{ti}^2\}.$$

For $(\alpha, \underset{\sim}{\beta}')' \in \Lambda_t$, define $a_i = \varepsilon(\alpha + x_i'\underset{\sim}{\beta})/(4b_t)$. As in the proof of Theorem 1, we have

$$P\{\sum_{i=1}^{n} |e_i| - \min_{\substack{(\alpha,\beta')'\in\Lambda_t \\ \ell\geq 1}} \sum_{i=1}^{n} |e_i - \ell a_i| \geq 0\}$$

$$\leq 4\exp\{-C\sum_{i=1}^{n} a_{ti}^2\}. \tag{5.4}$$

Noticing that $\varepsilon/(4b_t) \leq 1$, we have

$$\{(\ell\varepsilon/4b_t)(\alpha, \underset{\sim}{\beta}')': (\alpha, \underset{\sim}{\beta}')' \in \Lambda_t, \ \ell \geq 1\}$$

$$\supset \{\ell(\alpha, \underset{\sim}{\beta}')': (\alpha, \underset{\sim}{\beta}')' \in \Lambda_t, \ \ell \geq 1\}. \tag{5.5}$$

Hence, by (5.4) and (5.5) we obtain

$$P\{\sum_{i=1}^{n} |e_i| - \min_{\substack{(\alpha,\beta')'\in\Lambda_t \\ \ell\geq 1}} \sum_{i=1}^{n} |e_i - \ell(\alpha + x_i'\underset{\sim}{\beta})| \geq 0\}$$

$$\leq 4\exp\{-C\sum_{i=1}^{n} a_{ti}^2\}. \tag{5.6}$$

Since

$$\sum_{i=1}^{n} a_{ti} = (\varepsilon/4b_t)^2 \sum_{i=1}^{n} (\alpha_t + x_i'\beta_t)^2$$

$$\geq \varepsilon^2/(16b_t^2)\,\beta_t' T_n \beta_t \geq \varepsilon\rho_n/(16d_n^2),$$

we have, by (5.6),

$$P\{\sum_{i=1}^{n} |e_i| - \min_{\substack{(\alpha,\beta')'\varepsilon\Lambda_t \\ \ell\geq 1}} \sum_{i=1}^{n} |e_i - \ell(\alpha + x_i'\beta)| \geq 0\}$$

$$\leq 4\exp\{-C\rho_n/d_n^2\}, \tag{5.7}$$

where $C > 0$ is a constant independent of n, but may be taken as different value at each appearance.

From (5.3) and (5.7), we finally obtain that

$$P\{\sum_{i=1}^{n} |e_i| - \min_{\alpha^2+\beta'\beta\geq\varepsilon^2} \sum_{i=1}^{n} |e_i - \alpha - x_i'\beta| \geq 0\}$$

$$\leq 4((p+1)n)^{(p+1)k}\exp\{-C\rho_n/d_n^2\}. \tag{5.8}$$

In view of condition (2.1), applying Borel–Cantelli Lemma we get

$$\hat{\alpha}_n \to 0, \qquad \text{a. s.}$$

$$\hat{\underset{\sim}{\beta}}_n \to 0, \qquad \text{a. s.}$$

These prove the strong consistency of $\hat{\alpha}_n$ and $\hat{\underset{\sim}{\beta}}_n$. The last assertion of the theorem follows from (2.5) and (5.8). The proof is complete.

REFERENCES

1. Hoeffiding, W. (1963). J. Amer. Statist Assoc. **58** No.3, 13-30

2. Huber, P. J. (1981). Robust statistics. John Wiley & Sons, INC, New York

3. Li, K. C. (1984). Regression Models with Infinitely Many Parameters: Consistency of Bounded Linear Functionals. Ann. Statist **12**, 601-611

4. Oberhofer, W. (1982). The consistency of nonlinear regression minimizing the L_1-norm. Ann. Statist **10** No.1, 316-319

END

12-87

DTIC